Plum Creek

September 12, 2013

Gene Foster
TMDL manager
DEQ Water Quality Section

CC: Zach Loboy, David Waltz, Ryan Michie, Shannon Hubler, Josh Seeds, Kami Ellingson, Maryanne Reiter, Greg Peterson, Chris Jarmer, Randy Hereford, Jim James

RE: Comments and questions about use of aquatic macroinvertebrates for identifying water quality impairments and for setting improvement targets

Dear Gene:

I am submitting these comments on behalf of industrial forest landowners. After reviewing the DEQ material on use of aquatic macroinvertebrates to assess water quality conditions, plus source material on the topic, we remain concerned that this proposed approach does not provide a reliable basis for 303D listings or for recommending changes to land use practices. The approach may be useful for trend monitoring, or as a research tool to generate hypotheses regarding sediment or temperature conditions and trends in a watershed, but the cause-effect inferences made using this approach are too poorly established for it to be useful in pinpointing water quality impairments in a nonpoint realm. The indices DEQ has created (PREDATOR, FSS, and TS) and their associated statistics are far too removed from the ecology of the streams they are trying to describe, leaving us, as stakeholders, unconvinced that we have a method that can detect a management signal from natural background noise with any reasonable amount of certainty. How can DEQ move forward with developing biocriteria when the foundation of the approaches they use – natural conditions – are so poorly understood? As we've said numerous times during the mid-coast TMDL process, 'getting it right' when defining the problem is a precursor to productive problem solving. We do not outright dismiss the potential value of aquatic macroinvertebrates as indicators of stream conditions; however we suggest the DEQ continue with research on variability of insect communities and their response to disturbance.

You have taken on a large scale predictive modeling process for a large region that supports diverse landscape and stream conditions and are attempting to apply it to individual waterbodies. We feel that the basis of the current modeling is not supported by the limited sampling and research conducted to date. This could lead to gross misinterpretation of stream conditions. The topic is ripe for research and needs peer review. We urge DEQ to partner with OSU researchers in the Watersheds Research Cooperative where cause-effect linkages between watershed management and in-stream biota are being rigorously quantified.

We attach with our comments a review we solicited from Dr. Peggy Wilzbach of Humboldt State University.

Setting Expectations – Understanding Natural Variability and Response to Disturbance

The ability to characterize the diversity and relative abundance of macroinvertebrates in streams with 'least disturbed' or "reference" conditions is central to the success of all RIVPACS-type WQ evaluation tools (Hawkins et al. 2010, Stoddard et al. 2006). Because stream systems and the biota that live in them are dynamic, it is essential to account for spatial and temporal variability. DEQ's approach addresses spatial variability to some degree, but includes no adjustment for temporal variability. We believe this greatly limits the utility of these methods for use in WQ assessments and TMDLs.

Spatial Variability

The data used for development of DEQ's "PREDATOR" model (Hubler 2008) were mostly from samples gathered during EPA's Environmental Monitoring and Assessment Program (EMAP) or Oregon Watershed Enhancement Board (OWEB) efforts to describe environmental conditions of streams statewide. A statistically valid sampling design (e.g., the Generalized Random Tessellation Stratified [GRTS] approach) was used for these efforts, and some 407 macro-invertebrate samples were collected from 1998-2004.

Drake (2004) developed a method for using reach- or site-level disturbance information gathered during EMAP sampling to identify sites that exhibit 'least disturbed' or reference conditions. He calculated that at least 270 reference sites would be needed to characterize the expected natural conditions of the state of Oregon. Hubler (2008) used this guidance to identify a subset of 205 reference sites from the available macroinvertebrate samples across Oregon. So far as we know, no further site selection or sampling was done explicitly for the purpose of establishing a reference site network. The Marine Western Coastal Forest region RIVPACS model ultimately was based on 38 reference sites, of which 28 were in the Coast Range[1] and 10 were in the Willamette Valley Ecoregions (Hubler 2008). Are 28 sites adequate to characterize the Coast Range? This answer depends on the patchiness of aquatic macroinvertebrate assemblages and the homogeneity of streams in the Coast Range.

Gebbler (2004) examined variability in aquatic invertebrate community metrics for nine reaches in an Arizona stream. He concluded that "larger sample sizes (tens to hundreds) are necessary to obtain reasonable estimates of metrics and sample statistics". All of his reaches were from a "fairly homogenous" segment of the same river. Oregon coast range streams can have very different characteristics, especially for dominant substrates, owing to underlying lithology. Dr. Wilzbach points out in her comments that the DEQ's reference sites are disproportionately weighted toward drainages with relatively resistant lithologies. This may have produced a bias toward insect communities adapted to living in 'cleaner' environments. Such a bias might explain why Flynn Creek, an unmanaged basin that drains highly erodible sandstone, was initially judged to be sediment 'impaired' based on DEQ's biocriteria. The reference sites might

[1] Why does the Huff et al. (2006) report list 52 reference sites in the Coast Range Ecoregion, but the Hubler (2008) report lists only 28?

also have experienced different disturbance histories (e.g., floods, debris torrents, splash-dams, stream cleaning efforts) depending on where and when they were sampled[2]. We believe that 28 sites from the Coast Range are not enough to describe spatial variability.

Temporal Variability

Understanding variation of macroinvertebrate assemblages through time is just as important as understanding their spatial patchiness. Most of the samples used in the PREDATOR model and FSS were gathered during the period June-October, 1998-2004 (Huff et al. 2006, Hubler 2008). Aquatic macroinvertebrate abundance (and therefore capture probability) changes within and among years. Gravelle et al. (2009) tracked various metrics of aquatic insect communities for twelve years in a paired watershed study in Idaho. They found "substantial inter-annual variability" among years in the unmanaged control watershed (Figure 1). Li et al. (unpublished data) studied densities of benthic invertebrates in the Hinkle Creek Paired Watershed Study and also found substantial variability, both within and among years (Figure 2). Hawkins et al. (1982) relied on observed "distinct changes in taxonomic composition between seasons" to justify statistical independence of samples taken 2 months apart (June – October).

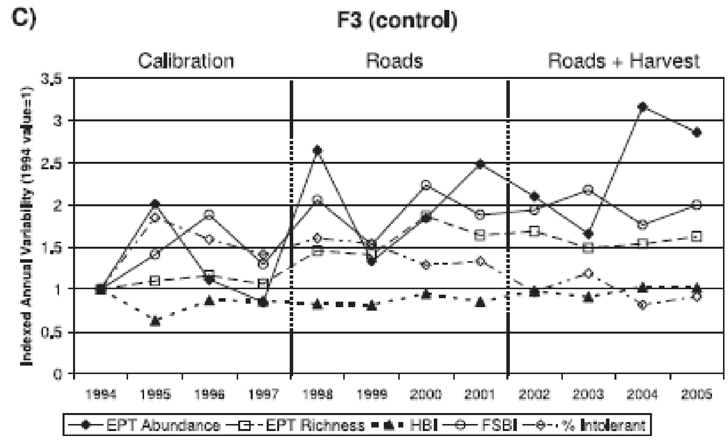


Figure 6. Examples of annual variability in metric values in the Mica Creek Experimental Watershed, 1994–2005. The Indexed Annual Variability values are indexed from 1994 values for comparison (1994 = 1.0), so 1.5 represents a 50% increase in metric value from the 1994 base value. (A) F1 (clearcut); (B) F2 (partial cut); and (C) F3 (control).

Figure 1. From Gravelle et al. 2009. FOR. SCI. 55(4):352–366

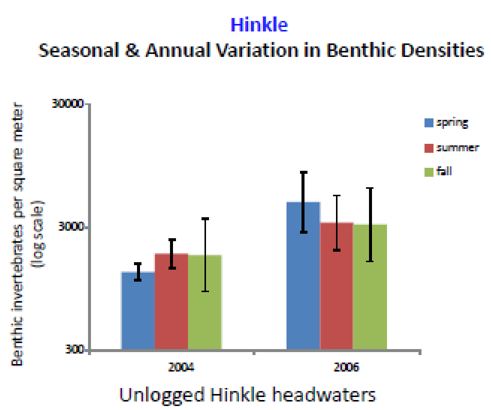


Figure 2. From Li et al. (unpublished). Long-term studies of macroinvertebrate response to harvest in Hinkle, Alsea and Trask watersheds. Presentation made at the Oregon State Univ. Paired Watersheds conference, April 18, 2013. http://wrcpairedwatershed2013.com/

In the May 28, 2013 memo to the mid-coast TMDL sediment TWG, DEQ reported the inter-annual variation in Fine Sediment Scores (FSS) was about 3 for the statewide subset of sites with more than 1 year of sampling. This seems low, given the inter-annual abundance differences reported in the Oregon and Idaho studies, and the way the FSS is calculated (sum of relative abundances * optima for selected taxa). Even if this variation accurately reflects temporal variability in FSS, a difference of 3 could dramatically change the interpretation of the sediment levels for the Coast Range Ecoregion (median of 6, 75th percentile impairment benchmark of 9; Huff et al. 2006, Table 4). The inter-annual variation (27%) was much higher for values of Observed/Expected

[2] For instance, streams sampled in 1998 after the 1996 extreme precipitation event might have a different sediment signature than ones sampled in 2004.

ratio derived from the PREDATOR model. A water body assessed with this index could therefore be judged as 'impaired' one year and 'unimpaired' the next, simply because of natural changes in taxa abundances.

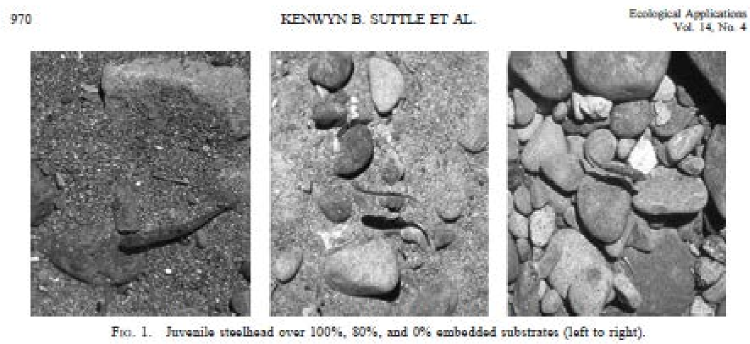
When preparing for the Alsea Watershed Study - Revisited (part of the Watersheds Research Cooperative), we requested that DEQ conduct a pre-treatment macroinvertebrate survey of Needle Branch in conjunction with planned monitoring at the Flynn Creek reference site. DEQ staff reported they were unable to comply because flows were so low in Needle Branch and protocols were not adapted to these conditions. Needle Branch, a stream that supports coastal Coho salmon, is discontinuously perennial, with reaches that go subsurface some years and reaches that always have surface flow. These conditions are not unusual in the Coast Range and they demonstrate how both spatial and temporal (within and between years) variability can affect results. If this natural variability is not accounted for then interpretations of conditions are not supportable.

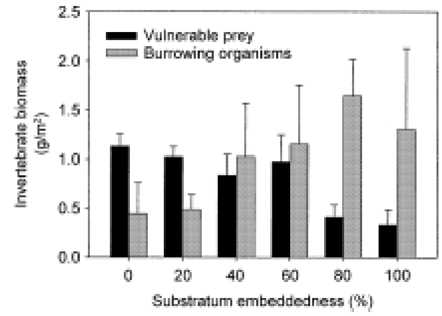
Hubler (2008) recommends repeated sampling at reference sites to establish whether O/E scores change over time. We concur, and ask how the DEQ intends to incorporate temporal variability in the O/E and FSS approaches? We recommend that DEQ re-characterize the richness and abundance of Coast Range macroinvertebrates using a statistically robust sampling scheme tailored to this area and then re-sample the reference sites for several years before attempting to develop and use PREDATOR and FSS as reliable determinants of water quality conditions.

Quantification of Specific Stressors – Fine Sediment Score (FSS)

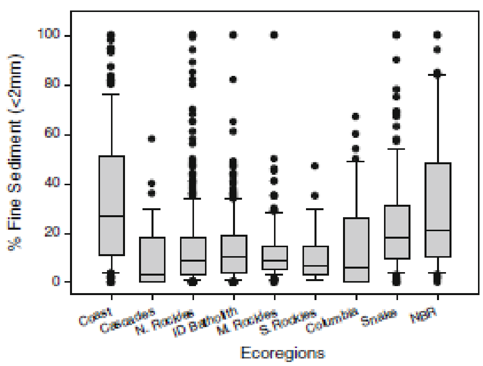
The same issues that concern us for development of the PREDATOR model - representative sampling, sample size, and characterization of natural background variability – weaken DEQ's effort to attribute shifts in macro-invertebrate assemblage composition to sediment or temperature. This approach also introduces potential errors from statistical inference techniques and from field methods. To begin with, Huff's (2006) weighted average inference model is based on the premise that "an ecologically sound estimate of a taxon's optima is … the mean value for all the sites where it is found, weighted by its log transformed abundance at each site" (Huff et al. 2006) or that "a taxon's tolerance is one standard deviation from the mean of the value, weighed by the taxon's log transformed abundance (Birks et al. 1990). We question this premise, owing to a lack of direct, quantitative measurements of taxa response to fine sediment to support it, and the unrealistically low levels of fines predicted for reference sites in the Coast Range Ecoregion.

How sensitive are macrobenthos to fine sediment levels? In the literature we could find, they don't appear to be nearly as sensitive as the FSS suggests. For example, Suttle et al. (2004) measured fish and invertebrate responses to a wide range of fine sediment levels (<2mm). They found no changes with fines up to 20% (see photo and graph).





Hawkins et al (1982) attempted to demonstrate sensitivity of stream macroinvertebrates to sediment. They associated community assemblages with the percentage sand (~1-mm) in stream substrates and they manipulated the amount of sand (0 to 100%) in trays of substrate they placed in the same streams. They found "substrate composition was seldom important, affecting the abundances only of shredders and filter feeders. Trays with higher amounts of fine substrate had had higher abundances of animals." Also, food quality was more important than substrate composition for its effect on abundance of different macroinvertebrate guilds.



Relyea et al. (2012), in their analysis of fine sediment (<2-mm) data gathered from western U.S. EMAP sites found no taxa's 75th percentile of occurrence was in streams of less than 10% fine sediment. This included the Coast Range Ecoregion (Washington and Oregon) which had the highest median percentage fine sediment (27%) (see figure at left)

These responses conflict with inferred sensitivities of the FSS method where average fine sediment difference between 'optimal' and 'impaired' for the list of top 30 taxa[3] used to develop the FSS (Huff et al. 2006, Table 2) was only 3.5% (±0.5% SD). Such as small number has to be well within observer error and seems far below any reasonable levels that would induce a biological response. Finally, the R^2 value for the non-metric multidimensional scaling method of associating sediment and aquatic insect assemblage structure was only 0.35 (Huff et al. 2006). Although this may be statistically significant, it may not be ecologically significant.

Part of the reason for the low R^2 value could be that the method of calculating the FSS is biased toward high fine sediment outcomes. By multiplying the relative abundance of taxa by their inferred sediment optima, then summing this for a sample, the presence of just a few taxa with a high fine sediment tolerance will produce a high FSS. This could happen without there actually being high fine sediment levels in the sampled stream. Sediment is not inherently 'bad', and the species that are associated with it are not 'bad'. They are members of the natural macroinvertebrate community. Streams within basins dominated by sandstone lithologies will naturally have higher fine sediment loads than streams with hardrock lithologies yet both can support healthy salmon populations.

[3] As ranked by Pearson's absolute r.

The authors of the FSS development report (Huff et al. 2006) acknowledge a dearth of quantitative measures of macrobenthic invertebrates response to fine sediment: "…no studies could be located that provided a direct comparison with stream macroinvertebrates and temperature or fine sediment" (p. 22). Fortunately, research completed since Huff et al. (2006) does provide direct information on this topic (Gravelle et al. 2009, Li et al. unpublished data from Hinkle, Trask, and Alsea paired watershed studies). We recommend the DEQ apply the PREDATOR and FSS methods to the macroinvertebrate data from these studies to investigate the sensitivity of macroinvertebrates to the magnitude, duration, and frequency of sediment fluxes that are likely to occur naturally and in response to contemporary forest management activities.

Fine Sediment Field Measurements

The low R^2 values associating sediment and aquatic insect assemblage structure (Huff et al. 2006) may be partly due to the accuracy of methods used to measure fine sediment <0.06 mm (i.e., silt). Stream substrate composition measurements were made using a modified Wolman pebble count method (Peck et al. 2006). As Bunte et al. (2009) points out, there is inconsistency in how this method is used and what results are generated. The EMAP method yields results that are weighted (40% of samples) to conditions at the water's edge where silt is likely in greater abundance than the center of the channel where macrobenthos are sampled. Measurements were also only taken for wetted widths, so they don't reflect the system's sediment load per standard fluvial geomorphic methods. Finally, the measurement method itself – identifying the particles at the end of a pointed rod hung at fixed intervals along a transect – could be biased. Silt particles are so fine you can't feel it (Bunte et al. 2009), and if it's so small you can't see it among larger particles, then observers are left to infer its presence. This is neither reliable nor reproducible, and is the reason that Relyea et al. (2012) chose a 2-mm fine sediment particle size threshold to associate with macrobenthos: "because most stream monitoring protocols use some form of a Wolman pebble count and we consider 2 mm the smallest size one can measure reliably using pebble count methods." For these reasons, we don't have high confidence in the correlation between measured amounts of silt and aquatic insect response. It might be coincidental. This doesn't give us confidence in the ability of the method to discern normal from excessive amounts of fines.

<u>Temperature Score</u>

We did not examine this metric in detail, but our general concerns regarding natural variability and discerning management signals from background noise apply. We note that the mean and median 7-day maximum temperatures recorded for the reference sites examined in Huff et al. (2006) were above the 16°C numeric criterion for core cold water rearing areas. This illustrates our point that 'stress' in the form of temperature (and fine sediment) is a natural feature of stream systems. The key is to identify when the magnitude, frequency, and duration of these natural features have been shifted beyond the tolerances of native communities <u>and</u> are actually due to land management practices.

Benchmarks of Biological Condition

Issues surrounding selection of thresholds for index values are related to those discussed above for natural variability – you need to know the latter to do a credible job of the former. DEQ chooses to infer WQ impairments based on observed taxa richness less than 85% of Expected for the PREDATOR-model and on the 75th percentile for FSS and CART. These limits seem arbitrary. From Figure 2 in Hubler (2008), it appears that there are at most 22 taxa expected in the Marine Western Coastal Forest region used for the PREDATOR evaluation of mid-coast streams. A difference of 15% (4 taxa) would be enough to shift a site rating from "least disturbed" to "most disturbed". All of the reference sites were by definition "least disturbed." How can a site within the range of values found in these least disturbed sites be considered disturbed? Also, the taxa expected but not found in samples are assumed to be "lost". Where did they go? We think it likely they are not lost, but rather hidden in the cloud of natural and seasonal variability. The bounds on this cloud must be determined before DEQ can attribute differences among samples to "disturbance." It would be better to base thresholds on the measured behavior of aquatic insect community responses to differing types and degrees of true disturbance. Once the effect of fine sediment or other environmental variables on the distribution of taxa in reference sites is known, shifts in this distribution can be used to define meaningful ecological changes. This approach was encouraged by Fore et al. (1996):

> *"In general, setting scoring criteria as a percentage of a reference site is a poor approach because it fails to recognize that reference condition is more meaningful when defined as a range of values rather than a single value."*

Benchmarks derived from subgroups of reference samples (i.e., the CART assessment) suffer from a potential compounding of errors: (1) reference sites not sampled adequately in time and space (2) Fine Sediment Scores not accurately reflecting macroinvertebrate community sensitivity to sediment (3) fine sediment sampling errors, and (4) small group sample sizes.

Interpretation of Results– the Mid Coast Basins TMDL

Assuming the DEQ's biocriteria are acceptable, how does DEQ intend to interpret the status of the mid-coast basins using DEQ's current biocriteria assessment methods? The original impairment listings based on biological criteria (2010) or miscellaneous sedimentation information (1998) resulted in 303D listings of the main trunks of rivers and tributaries. These samples were haphazardly gathered, so naturally they couldn't be judged to characterize the entire mid-coast area. The 2012 samples sites were selected more rigorously using a GRTS approach (per Ryan Michie, June 19, 2013 sediment TWG meeting minutes) and thus were better able to characterize the WQ conditions in the mid coast. Results indicate that virtually every sampled tributary in the upper Siuslaw is not impaired. What does this suggest to DEQ? That the sediment-producing tributaries sourcing the "impaired" mainstem haven't yet been sampled? That the tributaries are likely not impaired, but the mainstem is? That the mainstem is behaving differently than the tributaries? That different flow-regimes (mainstem versus tributaries) have different biological assemblages and that the biocriteria may not accurately interpret sediment conditions in mainstems? That mainstem reaches naturally have finer sediment? Again, impairment criteria still suffer from the weaknesses of their foundation – i.e., a limited number of non-randomly sampled reference sites.

We're encouraged that the FSS method did not result in Flynn Creek being listed as sediment impaired in 2012[4], but we're still not confident the method is working reliably. How can we be confident that sampling next year won't result in an impairment listing without any material change in watershed condition?

<u>Summary</u>:

Our chief concern is that reference conditions in the Oregon coast range have not been adequately characterized for their spatial and temporal variability. All of DEQ's metrics rely on comparisons of macroinvertebrate taxa richness and abundance in reference sites with those in test sites.

Macroinvertebrates do have a use in water quality management – there have been bonafide cases where macrobenthic communities have responded to pollutants, but these cases involved severe water quality impairments with point source inputs (e.g., sewage outfall [Ohio EPA 1988] or toxic pollutants from nonpoint sources (acid mine drainage [Larsen et al. 1996]). In forested systems, sediment and heat energy are natural watershed inputs. The magnitude, duration, and frequency of inputs vary greatly in time and space. Stream-dwelling biota have evolved to persist in these variable environments. The poorly-defined effects of inputs, such as sediment and temperature, on macroinvertebrate communities in natural systems with high natural background variability confound the ability to develop a good nonpoint bio-assessment tool from macrobenthos. DEQ is right to explore WQ indicators like benthic macroinvertebrate community structure. However this method is not yet mature enough nor peer reviewed to use for WQ impairment listings or TMDL development. It has the potential to misinterpret impaired or least-impaired conditions. Rather it should be investigated further and in conjunction with carefully controlled watershed-scale experiments where insect responses to forest management and other land uses can be isolated and quantified.

Another concern we have is that DEQ has relied greatly on statistical analyses of data that weren't collected for development of biocriteria and probably are inappropriate for this use. The linking of analyses can compound errors and ultimately obscure the basic biology of macrobenthic invertebrates and their response to environmental variables. We think it better to use simple metrics that are tied tightly to more obvious, coarse-level responses that can be corroborated with other evidence. Fore et al. (1996) summarized this concern well:

> *"Tests of significance are overused by ecologists (Yoccoz 1991); they focus on detection of impacts rather than on their magnitude or importance (Stewart-Oaten 1996)*

We believe it is prudent for DEQ to suspend further biocriteria analyses until data have been collected using a sampling design suited for the use, and that characterizes natural variability in time and space.

---

[4] And that tributaries to Cummins Creek and Drift Creek wilderness areas were also judged to be unimpaired.

In future TMDLs, DEQ should contact stakeholders years in advance to develop a sampling plan and approach that works for the specifics of the watershed. This would allow for field data to be collected and analyzed collaboratively. This is not a tactic to stall development of useful predictive methods but a need for reliable and robust methods that everyone can agree upon

Recommendations

DEQ has a lot riding on these approaches and methods, and therefore so do landowners. We recommend the following:

1. Gather the data needed to better understand:
    a. Aquatic insect community characteristics in the Oregon Coast Range
        - ✓ Spatial and temporal variability in taxa richness and abundance, and associated O/E indices
        - ✓ Resample reference sites periodically using random sampling frame designed specifically for the Oregon Coast Range Ecoregion.
            - Expand universe of reference sites to increase sample size and to better represent under-sampled strata (re-examine criteria used to select reference sites)
    a. Response to disturbance – proximal vs. distal
        - ✓ Note that Li et al. (unpublished data) found that aquatic insect assemblages responded to proximal disturbances, not those upstream. Factor this into assessment methodology.
        - ✓ Measure behavior of FSS and O/E metrics in response to changes in bedded sediment (sand and smaller) as measured through conventional fluvial geomorphic methods. Validate indices by demonstrating they reflect changes in system condition.
    b. Response to disturbance – lag times
        - ✓ Samples are assumed to reflect current/instantaneous conditions of the water body from the point of sampling and areas upstream. Transport of sediments is not instantaneous. How is time lag addressed in the biocriteria sampling and TMDL process?
2. Revise the approach to assigning impairment levels to the range of taxa richness and abundance gathered from reference sites. Use a standard deviation from the mean instead of an arbitrary % of reference maximum. Include a temporal dimension.
3. Use a minimum detectable effect approach rather than setting benchmarks as % of reference sites.
4. Develop a weight-of-evidence procedure for linking benthic invertebrate samples with stream and upland conditions. Cissel et al. (2012) recently completed a Geomorphic Road Analysis and Inventory Package (GRAIP) analysis of sediment sources from the road system in the NF Siuslaw. This system was listed as impaired for sediment, and would be a good location to test these ideas.
    a. Perform field inspections of site and watershed conditions where aquatic insect info suggests impairment. Validate the ability of the indices to accurately characterize WQ.

   b. Compare FSS with FSBI (Relyea et al. 2012) for the Oregon Coast Range Ecoregion
   c. Evaluate fish populations and fish habitat in sampled basins.
      - ✓ Work with the watersheds research cooperative to establish cause-effect links between watershed management, water quality, and beneficial uses.
5. Peer review the results from recommendations 1-4, with emphasis on use of these methods for water quality impairment listings and TMDLs.
6. For future TMDLs, work with stakeholders years in advance of the TMDL to gather necessary data and to agree on approaches and methods.
7. Figure out a sampling scheme that will provide a picture of WQ conditions in the TMDL area, and develop a reliable approach to investigate sources where weight of evidence suggests a true problem exists.

Thank you for the continued dialogue and your willingness to make the TMDL process work for forest land owners. We appreciate DEQ’s cooperative approach to maintain and improve water quality in Oregon.

Sincerely,

Jeff Light
Forest Hydrologist
Plum Creek Timber Company
Jeff.light@plumcreek.com
541-270-1303

Literature Cited

Birks, H.J.B., J.M. Line, S. Juggins, A.C. Stevenson, and C.J.F ter Braak. 1990. Diatoms and pH reconstruction. Philosophical Transactions of the Royal Society of London. B. Series Biological Sciences 327: 263-278.

Bunte, K., S.R. Abt, J.P. Potyondy, and K. W. Swingle. 2009. Comparison of Three Pebble Count Protocols (EMAP, PIBO, and SFT) in Two Mountain Gravel-Bed Streams. J. Am. Water Res. Assoc. 45(5): 1209-1227.

Cissel, R., T. Black, N. Nelson and C. Luce. 2012. North Fork Siuslaw River GRAIP Watershed RODs Assessment, Siuslaw River Subbasin, Siuslaw National Forest, Oregon, US Forest Service, Rocky Mountain Research Station, Boise, Idaho, 67 p.

http://www.fs.fed.us/GRAIP/downloads/case_studies/WatershedStudies.shtml

Collins, G. 1995. Environmental Monitoring and Assessment Program: Methods Format Guidance. EPA 620/R-95/001. Environmental Monitoring and Assessment Program: U.S. Environmental Protection Agency, Cincinnati, Ohio.

Fore, L.S., J.R. Karr, and R.W. Wisseman. 1996. Assessing Invertebrate Responses to Human Activities: Evaluating Alternative Approaches. J. N. Am Benthol. Soc. 15(2): 212-231.

Gebbler, J.B. 2004. Mesoscale spatial variability of selected aquatic invertebrate community metrics from a minimally impaired stream segment. J. N. Am. Benthol. Soc. 23(3): 616:633.

Gravelle, J.A., T.E. Link, J.R. Broglio, and J.H. Braatne. 2009. Effects of timber harvest on aquatic macroinvertebrate community composition in a northern Idaho watershed. For. Sci. 55(4):352–366.

Hawkins, C.P., R.H. Norris, J.N. Hogue, and J.W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10(5): 1456-1477.

Hawkins, C.P., M. L. Murphy, and N.H. Anderson. 1982. Effects of canopy, substrate composition, and gradient on the structure of macroinvertebrate communities in Cascades Range streams of Oregon. Ecology 63(6): 1840-1856.

Hubler, S. L. 2008. PREDATOR: Development and use of RIVPACS-type macroinvertebrate models to assess the biotic condition of wadeable Oregon streams. Oregon Dept. Envr. Qual. 51 pp.

Huff, D.D., S.L. Hubler, Y. Pan, and D.L. Drake. 2006. Detecting shifts in macroinvertebrate assemblage requirements: Implicating causes of impairment in streams. Oregon Dept. Envr. Qual. 37 pp.

Kaufmann, P.R., P. Levine, E.G. Robison, C. Seeliger, and D.V. Peck. 1999. Quantifying physical habitat in wadeable streams. EPA 620/R-99/003. Environmental Monitoring and Assessment Program, U.S. Environmental Protection Agency, Corvallis, OR.

Peck, D.V., A.T. Herlihy, B.H. Hill, R.M. Hughes, P.R. Kaufmann, D.J. Klemm, J.M. Lazorchak, F.H. McCormick, S.A. Peterson, P.L. Ringold, T. Magee, and M. Cappaert, 2006. Environmental

Monitoring and Assessment Program-Surface Waters Western Pilot Study: Field Operations Manual for Wadeable Streams. EPA⁄620⁄R-06⁄003. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.

Relyea, C.D., G.W. Minshall, and R.J. Danehy. 2012. Development and validation of an aquatic fine sediment biotic index. Environmental Management 49(1): 242-252.

Stewart-Oaten, A. 1996. Problems in the analysis of environmental monitoring data. Pp. 109-131 in: *Detection of Ecological Impacts: Conceptual Issues and Application in Coastal Marine Habitats*, R.J. Schmitt, and C.W. Osenberg, eds. Academic Press, San Diego, CA.

Stoddard, J. L., D. V. Peck, S. G. Paulsen, J. Van Sickle, C. P. Hawkins, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. P. Larsen, G. Lomnicky, A. R. Olsen, S. A. Peterson, P. L. Ringold, and T. R. Whittier. 2005. *An Ecological Assessment of Western Streams and Rivers*. EPA 620/R-05/005, U.S. Environmental Protection Agency, Washington, DC.

Yoccoz, N.G. 1991. Use, overuse and misuse of significance tests in evolutionary biology and ecology. Bull. Ecol. Soc. Am. 76: 106-111.

CALIFORNIA COOPERATIVE FISH & WILDLIFE RESEARCH UNIT
U. S. Geological Survey
Humboldt State University
1 Harpst Street, Arcata, CA 95521
707-826-3268   FAX: 707-826-3269
E-Mail: cuca@humboldt.edu

June 27, 2013

Memorandum

TO: Jeff Light, Plum Creek Timber

FROM: Peggy Wilzbach

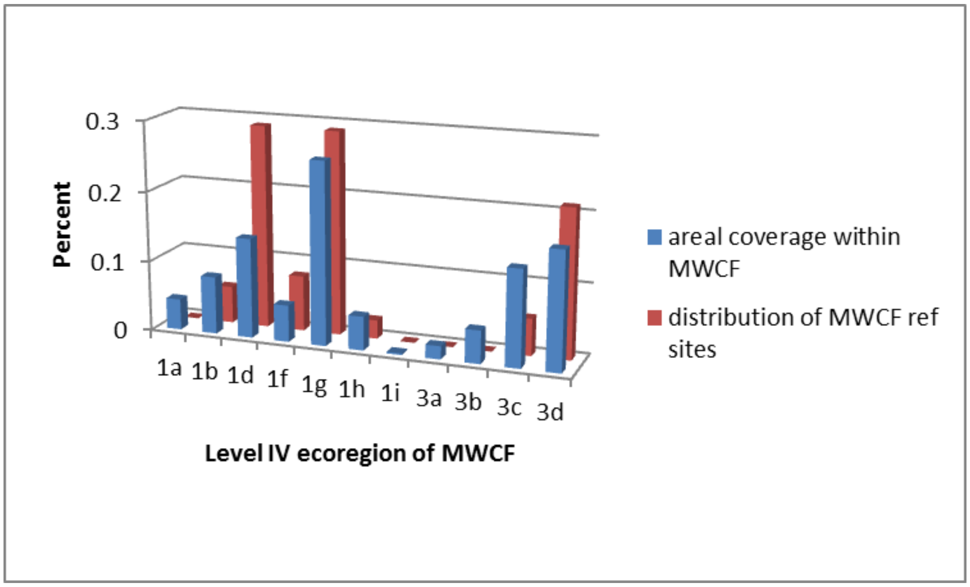
SUBJECT: Comments on DEQ predator model

As we discussed on the phone, the various bioassessment methods each have some strengths and limitations. I have not been a huge fan of RivPacs for a number of reasons. First, its reliance on taxonomic identification makes it inaccessible to most lay groups, and the ability to even resolve taxonomic identity is quite uneven among invertebrate groups (e.g. while mayflies are readily taken to species, few people in this country are able to identify chironomids at a resolution higher than subfamily or tribe). Years can lag between sample collection and procurement of data, becauses samples back up in the few labs equipped to do taxonomic identification. It concerns me that samples are collected only from riffle habitats in RivPacs protocols (this limitation is also common to invertebrate IBI's), as this limits the amount of information that can be obtained, and limits as well the use of the approach in rivers lacking a riffle-pool morphology. My biggest concern with Rivpacs, however, is that the relationship between invertebrate taxonomic composition and ecosystem functioning or health is not certain. The presence or absence of specific taxa reveals little about critical ecosystem functions including decomposition, nutrient cycling, and secondary production. Does it matter to a hungry fish if a stream harbors *Baetis bicaudatus* but not *Baetis tricaudatus*? Do the two species differently impact primary production and algal assemblages, or leaf decomposition? I dunno...perhaps, but our understanding of stream ecology is not generally advanced enough to predict how. Information that the ratio of observed to expected taxa differs from that found in a reference site doesn't by itself point the way toward either a ready diagnosis of potential problems or suggest a prescription for recovery.

My own objections aside, however, RivPacs is nonetheless a legitimate approach, and one that has been adopted by many state agencies, including Oregon DEQ. So what follows are just a few comments and questions I have about DEQ's reference site selection and Predator model development, based on the two OR DEQ technical reports (DEQ08-LAB-0048-TR and TR WAS04-002) and the datasets (output_bugs_CART_fianl_20121008 and Raw bug data_CR only_5.30.13) that you shared with me. I've focused my attention on just the model developed for the Marine Western Coastal forest (MWCF).

Reference Site Selection: there are 2 major issues to consider here. The first concerns sample size and sampling frequency of reference sites; the second and more important issue is whether environmental variability is accounted for in the selection of reference sites.

- The MWCF model includes 38 reference sites. Is this number adequate? It's hard to tell.   The number of reference sites required for bioassessment should increase with landscape heterogeneity (Yoder and Rankin 1995), and considerable environmental heterogeneity occurs within the MWCF. The MWCF subsumes 11 different level IV ecoregions (7 in the Coast Range, 4 in the Willamette Valley), and this heterogeneity is expressed in varying physiography, elevation, geology, soils, climate, and vegetation. DEQ states that its goal is to obtain 10-20 reference sites per region/gradient grouping, but neither of the documents describe which if any primary natural gradient(s) were chosen as a grouping factor, and how many levels were chosen within a gradient.
- Questions about sampling frequency are related to sample size issues. The Reference Site Selection document mentions the intent to re-sample annually a subset of reference sites to evaluate whether shifts have occurred in reference site conditions. The dataset, which covers 1984-2004, gives no indication that re-sampling has occurred. Temporal changes in composition of invertebrates, of course, would be likely to accompany successional changes that occur in stream settings. Even in completely pristine systems, I wouldn't be confident that invertebrate assemblage structure would remain constant after a lapse of 9-26 yrs. And given marked changes in global climate that have occurred since the mid-1980s, climatic and perhaps vegetative changes are likely to have occurred that are broader in scope and scale than localized anthropogenic effects.
- Are selected sites representative of the range of environmental conditions occurring within the region? One way of looking at this is to compare the areal extent of each of the level IV ecoregions within the MWCF (14,395 sq mi) with the distribution of the 38 reference sites. What is immediately striking is that the Coast Range ecoregion 1d [Volcanic] is over-represented in references sites relative to its areal extent within the MWCF, comprising 14 % of the land area but 29% of all reference sites. Especially under-represented among reference sites is the Willamette Valley ecoregion 3c (Prairie Terraces), which also comprises 14% of the MWCF area but makes up only 5% of the reference sites (see below).



- Not surprisingly, the Volcanics ecoregion takes its name from its volcanic geology of extrusive igneous rock, incuding basalt flows and concreted basalt materials. While this ecoregion experiences a high potential for landslides as a result of abundant precipitation, steep slopes and high uplift rates, the rock is also more resistant to erosion than is sedimentary rock, and I wouldn't expect streams in the ecoregion to persistently carry high loads of suspended sediments. On the other hand, the Prairie Terraces ecoregion derives from fluvial deposits from the Missoula floods, and the low gradient streams in this area are deeply entrenched within banks of clay. Even though erosion rate is low, streams are often turbid, irrespective of anthropogenic influence. With an under-representation among reference sites of naturally turbid streams and over-representation of streams underlain by resistant rock, taxonomic composition of invertebrate assemblages may be slanted toward taxa without adaptations for dealing with fine sediments.

Model Development: the Predator model developed by DEQ follows well-established protocols of the RIVPACS approach. My only quibbles here are concerned with the choice of invertebrate detection probabilities, and with the approach to developing and choosing predictor variables.

- Invertebrate taxonomy: The choice of a detection probability for invertebrate taxa of greater than 50% results in a fairly coarse level of taxonomic resolution (operational taxonomic units), which makes it more difficult to detect differences that do occur between reference and test sites. Curiously, this seems to run counter to the underlying philosophy of the Rivpacs approach that the 'devil is in the details' - i.e. that fine differences in taxonomic composition between reference and test sites can be used to assess stream health and suggest causes of impairment.
- Predictor variables: A decision to restrict predictor variables to those obtainable from GIS coverages may be expedient, but it's unfortunate in that it limits the information available to differentiate among reference groups. Riparian type, for example, or

substrate composition, are both known to strongly affect invertebrate assemblages, but these data are rarely available in GIS coverages.

- I understand that values of predictor variables measured at sites being assessed are tested to determine if they are within a statistically acceptable range of values measured at reference sites. A question I have that I didn't see addressed in the documents was whether data are testsed only with respect to the predictor variables used to build the model, or if all environmental data are included in this assessment. This would make a difference. For example, the MWCF model uses 2 predictor variables, of julian date and longitude. If a test site meets the date and longitude criteria, but test and reference sites differ in basin geology, would this be detected?
- I believe that Julian date is a poor choice as a predictor variable, as it doesn't identify any unique environmental condition around which invertebrate communities might be assembled. Many invertebrate taxa, particularly within orders of mayflies and true flies, have more than 1 generation per year. Other taxa have a univoltine or longer life history, but can be found in the water only during a certain and often limited season, as they may be in an egg or other resting stage. Periods of activity differ among taxa – some species grow primarily during temperate winters, and others grow primarily during spring and early summer. So the taxonomic composition of the invertebrate assemblage from a single site may appear to differ depending on the time of year in which a collection is made, even with sampling times restricted from May to October.